**System Development Foundation Benchmark Series**

Michael Brady, editor
*Robotics Science,* 1989

Max V. Mathews and John R. Pierce, editors
*Current Directions in Computer Music Research,* 1989

Philip R. Cohen, Jerry Morgan, and Martha E. Pollack, editors
*Intentions in Communication,* 1990

Eric L. Schwartz, editor
*Computational Neuroscience,* 1990

# Intentions in Communication

edited by Philip R. Cohen, Jerry Morgan, and Martha E. Pollack

System Development Foundation Benchmark Series

A Bradford Book
The MIT Press
Cambridge, Massachusetts
London, England

# Chapter 2

# What Is Intention?

*Michael E. Bratman*

## 1 *Introduction*

In a symposium on "Intentions and Plans in Communication and Discourse" it is only fitting for a philosopher to ask, What is intention? This is a question I have discussed at some length in a recent book (Bratman 1987). Here my aim is to sketch in condensed form some of the main ideas in that book, focusing on those that seem to me to be particularly relevant to interdisciplinary research on the nature of intelligent agency.

Begin by making a distinction (see Anscombe 1963). We use the concept of intention to characterize both our actions and our minds. You believe (correctly) that I have *intentionally* written this sentence and that I have written it *with the intention* of illustrating a certain distinction. Perhaps you also believe that when I wrote this sentence in January I *intended* to discuss it at the symposium in March. In the former case you use the concept of intention to characterize my action; in the latter case you use the concept of intention to characterize my mind.

In approaching the question "What is intention?" I propose beginning with intending to do something. It is part of our commonsense conception of mind and action that we sometimes intend to do things and that such intentions typically are intentions to act in certain ways in the future. As a first step toward answering our general question about the nature of intention, I want to know what it is to have such a future-direct intention.

How do we go about answering such a question? My approach will be broadly speaking within the functionalist tradition in the philosophy of mind and action. We say what it is to intend to do something by specifying the "functional roles" characteristic of intention. To do this, we need to articulate the systematic relations between such intentions, other psycho-

I am indebted to many people for their help. Among them let me especially mention the members of the Rational Agency Project at CSLI, in particular: Martha Pollack, Charles Dresser, David Israel, Philip Cohen, Mark Crimmins, Todd Davies, Mike Georgeff, Pat Hayes, Kurt Konolige, and Amy Lansky. This research benefited from support from the Center for the Study of Language and Information, made possible in part through an award from the System Development Foundation.

logical states (such as belief and desire), relevant psychological processes and activities (such as practical reasoning), and crucial inputs and outputs: perception and action. Our commonsense understanding of intention, belief, desire, perception, and action depends on the supposition of appropriate underlying regularities within which these phenomena are embedded. These regularities will doubtless be what Grice (1974–75) calls "ceteris paribus laws" and in articulating them we will surely need to abstract away from many complexities. But reliance on ceteris paribus regularities and appropriate abstraction comes with the territory.

When we try in this way to specify the functional roles characteristic of intention, we are quickly faced with a trilemma. Suppose I intend today to drive over the Golden Gate bridge tomorrow. My intention today does not reach its ghostly hand over time and control my action tomorrow; that would be action at a distance. But my intention must *somehow* influence my later action; otherwise, why bother today to form an intention about tomorrow? It will be suggested that, once formed, my intention today will persist until tomorrow and then guide what will then be present action. But presumably such an intention is not irrevocable between today and tomorrow. Such irrevocability would clearly be irrational; after all, things change and we do not always correctly anticipate the future. But this suggests that tomorrow I should continue to intend to drive over the Golden Gate then only if it would be rational of me then to form such an intention from scratch. But then why should I bother deciding today what to do tomorrow? Why not just cross my bridges when I come to them? So it may seem that future-directed intentions will be either (1) metaphysically objectionable (since they involve action at a distance), or (2) rationally objectionable (since they are irrevocable), or (3) just a waste of time.

Such a trilemma might lead some to be suspicious of the very idea of future-directed intention and to suppose that a coherent model of intelligent agency should to without this idea. Indeed, I suspect that worries of the sort captured by this trilemma may go some way toward explaining the tendency in mid–twentieth century philosophy of action to ignore future-directed intention and to focus primarily on intentional action and action done with an intention. But I believe that future-directed intentions play a central role in our psychology, both individual and social, and that it is a serious error to ignore them in a theory of intelligent activity. My proposal, then, is to look for a functionalist treatment of future-directed intention, one that makes it clear how to avoid being impaled on one of the horns of our trilemma.

My procedure will be as follows. I will turn first to the role of future-directed intentions as inputs to ongoing practical reasoning and planning. This will lead me to certain questions about belief and about its relation to intention. In particular, I will focus on the distinction between what one intends and what one merely expects to result from what one intends. This discussion will lead us to a more complex view than we might otherwise have had of the relation between practical reasoning and the intentions in which it issues. At the end I will return to our trilemma to check to see that we have escaped its horns. Throughout I have the hope (dare I say the intention?) that the conception that emerges of intention and its role in our lives will not only seem plausible as a partial conception of human agency but also prove useful in developing a general but realistic model of intelligent agency.



## 2 *Intention and Practical Reasoning*

We frequently settle on an intention concerning the future and then proceed to reason about how to do what we intend. Today I decided to go to Monterey tomorrow. Now I must figure out how to get there. In such reasoning my future-directed intention to go to Monterey functions as an important input; and its role as such an input is, I believe, central to our understanding of intention. I want to say what this role is.

Now the philosophical literature about practical reasoning has tended to be somewhat schizophrenic. The dominant literature sees practical reasoning as beginning with certain of the agent's desires and beliefs and issuing in a decision, choice, or action. Practical reasoning is a matter of weighing conflicting considerations for and against conflicting options, where the relevant considerations are provided by what the agent desires/values/cares about and what the agent believes. Practical reasoning consists in weighing desire-belief reasons for and against conflicting courses of action. This is the picture of practical reasoning sketched by Donald Davidson in his influential series of essays (Davidson 1980; Bratman 1985), and it is the picture that seems to lie behind most work in so-called decision theory (see, for example, Jeffrey 1983). What it is important to notice about this model of practical reasoning is that it provides no distinctive role for an agent's future-directed intentions as inputs to such reasoning. All practical reasoning is a matter of weighing desire-belief reasons for action.

A contrasting literature—best represented by Hector-Neri Castañeda's book *Thinking and Doing* (Castañeda 1975; Bratman 1983)—has emphasized the role of prior intentions as inputs into practical reasoning. The paradigm here, broadly speaking, is reasoning from a prior intention and relevant beliefs to derivative intentions concerning means, preliminary steps, or more specific courses of action. My close advisors on matters of artificial intelligence—David Israel and Martha Pollack—have taught me that this is also the paradigm that dominates there as well, in the guise of the "planning problem."

I think it is clear that these two traditions are each focusing on a real and important phenomenon: we do engage in the weighing of reasons both for and against conflicting options, and we do reason from prior, future-directed intentions to further intentions concerning means and the like. But how are these two kinds of reasoning related? This is a question that needs to be answered by an account of the role of intentions as inputs to further reasoning.

As a first step, return to our question: Why bother with future-directed intentions anyway? Why not just cross our bridges when we come to them? I think there are two main answers. The first is that we are not frictionless deliberators. Deliberation is a process that takes time and uses other resources; this means that there are obvious limits to the extent of deliberation at the time of action. By settling on future-directed intentions, we allow present deliberation to shape later conduct, thereby extending the influence of deliberation and Reason on our lives. Second, and relatedly, we have pressing needs for corrdination. To achieve complex goals, I must coordinate my present and future activities. And I need also to coordinate my activities with yours. Furture-directed intentions help facilitate both intra- and interpersonal coordination.

How? Part of the answer comes from noting that future-directed intentions are typically elements of larger plans. My intention today to go to Monterey tomorrow helps coordinate my activities for this week, and my activities with yours, by entering into a larger plan of action—one that will eventually include specifications of how to get there and what to bring, and one that will be coordinated with, for example, my child-care plans and your plans for meeting me in Monterey. And it will do this in ways compatible with my resource limitations and in a way that extends the influence of today's deliberation to tomorrow's action.

Since talk of plans is rampant in work in artificial intelligence, I had better stop right away and make it clear how I am using this term. A first distinction that needs to be made is between plans as abstract structures and plans as mental states. When I speak here of plans, I have in mind a certain kind of mental state, not merely an abstract structure of a sort that can be represented, say, by some game-theoretical notation. A more colloquial usage for what I intend might be "having a plan." But even after saying this, there remains room for misunderstanding; for there seem to be two importantly different cases of having a plan. On the one hand, I might have only a kind of recipe: that is, I might know a procedure for achieving a certain end. In this sense I can have a plan for roasting lamb whether or not I actually intend to roast lamb. On the other hand, to have a plan to roast lamb is to be planning to roast it: it involves intending to roast it. It is the second kind of case that I intend when I speak of plans. Plans, as I shall understand them, are mental states involving an appropriate sort of commitment to action: I have a plan to *A* only if it is true of me that I plan to *A*.



The plans characteristic of a limited agent like me typically have two important features. First, my plans are typically *partial*. When I decide today to go to Monterey tomorrow, I do not settle all at once on a complete plan for tomorrow. Rather, I decide now to go to Monterey, and I leave till later deliberation about how to get there in ways consistent with my other plans. Second, my plans typically have a *hierarchical structure*. Plans concerning ends embed plans concerning means and preliminary steps; and more general intentions embed more specific ones. As a result, I may deliberate about parts of my plan while holding other parts fixed. I may hold fixed certain intended ends, while deliberating about means or preliminary steps.

The strategy of settling in advance on such partial, hierarchically structured plans, leaving more specific decisions till later, has a pragmatic rationale. On the one hand, we need to coordinate our activities both within our own lives and, socially, between lives. And we need to do this in ways compatible with out limited capacities to deliberate and process information. This argues for being planning creatures. On the other hand, the world changes in ways we are not in a position to anticipate; so highly detailed plans about the far future will often be of little use and not worth bothering with. Partial, hierarchically structured plans for the future provide our compromise solution. And with the partiality of plans go patterns of reasoning in which my prior intentions play an important role as inputs: reasoning from a prior intention to further intentions concerning means or preliminary steps. In such reasoning we fill in partial plans in ways required for them successfully to guide our conduct.

To understand such reasoning, we need to reflect on demands that, other things being equal, plans need,to satisfy to serve well their roles in coordination and in extending the influence of deliberation. First, there are *consistency constraints*. Plans need to be both internally consistent and consistent with the agent's beliefs. Roughly, it should be possible for my plans taken together to be successfully executed in a world in which my beliefs are true. Second, though partial, my plans need to be filled in to a certain extent as time goes by. My plans should be filled in with subplans concerning means, preliminary steps, and relatively specific courses of action, subplans at least as extensive as I believe is now required to do what I plan. Otherwise, they will suffer from *means-end incoherence*.

Associated with these two demands are two direct roles intentions and plans play as inputs in practical reasoning. First, given the demand for means-end coherence, prior intentions frequently *pose problems* for further deliberation. Given my intention to go to Monterey tomorrow, I need soon to fill in my plan with a specification of some means to getting there.

Second, given the need for consistency, prior intentions *constrain* further intentions. If I am already planning to leave my only car at home for Susan to use, then I cannot consistently solve my problem of how to get to Monterey by deciding to take my car.

Prior intentions, then, pose problems for deliberation, thereby establishing standards of *relevance* for options considered in deliberation. And they constrain solutions to these problems, providing a *filter of admissibility* for options. In these ways prior intentions and plans help make deliberation tractable for agents like us, agents with substantial resource limitations.

This gives us a natural model of the relation between two kinds of practical reasoning: the weighing of desire-belief reasons for and against conflicting options, and reasoning from a prior intention to derivative intentions concerning means and the like. Our model sees prior intentions as elements of partial plans, plans that provide a *background framework* within which the weighing of desire-belief reasons is to occur. It is this framework that poses problems for such further reasoning and constrains solutions to these problems. So practical reasoning has two levels: prior, partial plans pose problems and provide a filter on options that are potential solutions to these problems; desire-belief reasons enter as considerations to be weighed in deliberating between relevant and admissible options.

All this requires that prior intentions and plans have a certain *stability*: prior intentions and plans resist being reconsidered or abandoned. If we were constantly to be reconsidering the merits of our prior plans, they would be of little use in coordination and in helping us cope with our resource limitations. However, as noted earlier, it would also be irrational to treat one's prior plans as irrevocable, for the unexpected does happen. This suggests that rational agents like us need general policies and habits that govern the reconsideration of prior intentions and plans. Roughly speaking, their nonreconsideration should be treated as the "default," but a default that is overridable by certain special kinds of problems. An important area of research is to say more about what sorts of policies and habits concerning the reconsideration of prior plans it would be reasonable for limited agents like us to have. I believe that the consideration of such matters will lead us into general issues about aspects of rational agency that go beyond reasoning and calculation; but I cannot go into this matter further here.

So we now have a sketch of one of the major roles of future-directed intentions in the psychology of limited rational agents like us. As elements of partial, hierarchical plans, prior intentions pose problems and filter out options for deliberation. And throughout they need to exhibit an appropriate kind of stability, resisting reconsideration except when faced with certain unanticipated problems.



This account of the role of prior intention in further reasoning assumes that there is such a thing as flat-out belief. That is, it assumes that we sometimes just believe certain things—for example, that I have a meeting in Monterey tomorrow and that my home is not walking distance from Monterey—and do not just have degrees of confidence or "subjective probabilities" ranging from 0 to 1. Such flat-out beliefs combine with my plans to make certain options inadmissible. If I just assigned a high probability (but a probability less than 1) to my having only one car, the plan of driving a car of mine to Monterey while leaving a car of mine at home for Susan would not run into problems of inconsistency, strictly speaking. So the option of driving a car of mine to Monterey would not be inadmissible. It is my flat-out belief that I have only one car that combines with my prior plans to make inadmissible the option of driving a car of mine to Monterey.

The background framework against which practical reasoning proceeds includes, then, flat-out beliefs as well as prior intentions and plans. Of course, just as I can always stop and reconsider some prior intention, I can also stop and reconsider some background flat-out belief. Still, in a normal case in which there is no such reconsideration my planning will be framed in part by my relevant flat-out beliefs.

None of this denies the importance to practical reasoning of subjective probabilities less than 1. The idea, rather, is to see such subjective probabilities as entering into deliberation, but deliberation that is already framed by the agent's prior plans and flat-out beliefs.

### 3 *Intention and Belief*

I need now to discuss how intention is related to belief. This is a complex issue, and I can only discuss some elements of it here.

Our model of the role of future-directed intentions as inputs to further reasoning clearly depends on one main idea about the relation between intention and belief. This is the idea that there is a defeasible demand that one's intentions be consistent with one's beliefs. It is this demand that drives the operation of what I have called the filter of admissibility on options. But there is also a second idea implicit in the model. To see what this is, we need to return to the point—central to this discussion—that prior intentions and plans play a crucial coordinating role. And we need to say more about how that works. How does my prior intention to go to Monterey tomorrow help support coordination within my own life and, socially, between my life and yours?

I think the answer is that it does this normally by providing support for the expectation that I will in fact go there then. That is, my intention normally helps support your expectation that I will go there, thereby enabling you to plan on the assumption that I'll be there. And it normally

helps support my expectation that I will go, thereby enabling me to plan on a similar assumption. But how is it that my intention can provide such support?

Part of the answer has already been suggested. My intention will normally lead me to reason about how to get there and to settle on appropriate means. And my intention has a certain stability, so one can expect that it will not be easily changed, Still, there is a gap in the explanation. For all that we have said so far, my intention might stay around and lead to reasoning about means and yet still not control my conduct when the time comes.

What is needed here is a distinction between two kinds of pro-attitudes. Intentions, desires, and valuations are, and ordinary beliefs are not, *pro-attitudes*. Pro-attitudes in this very general sense play a motivational role: in concert with belief they can move us to act. But most pro-attitudes are merely potential influencers of conduct. For example, my desire to play basketball this afternoon is merely a potential influencer of my conduct this afternoon. It must vie with my other relevant desires—say, my desire to finish writing this paper—before it is settled what I will do. In contrast, once I intend to play basketball this afternoon, the matter is settled: I normally need not continue to weigh the relevant pros and cons. When the afternoon arrives, I will normally just proceed to execute my intention. My intention is a *conduct-controlling* pro-attitude, not merely a *potential influencer* of conduct.

It is because intentions are conduct controllers that the agent can normally be relied on try to execute them when the time comes, if the intention is still around and if the agent then knows what to do. It is because intentions are normally stable that they can normally be relied on to persist until the time of action. And it is because intentions drive relevant means-end reasoning and the like that the agent can normally be relied on to put herself in a position to execute the intention and to know what to do when the time comes. Taken together, these features of intention help explain how future-directed intentions normally support coordination by supporting associated expectations of their successful execution.

None of this entails, by the way, that *every* time you intend to $A$ in the future you must believe that you will $A$. That thesis seems to me a bit too strong, though I will not try to argue the point here. All that is required is that your intention will normally support such a belief, thereby supporting coordination.

So, future-directed intentions will normally be both consistent with the agent's beliefs and support beliefs in their successful execution. In the normal case, then, when I intend to $A$, I will also believe that I will $A$. Nevertheless, there remains an important distinction between intention and belief. I now want to focus on one aspect of this distinction: the distinction



between what one intends and what one merely expects as an upshot of what one intends. This will shed further light on the way in which intentions can be an output of practical reasoning.

## 4 *Intention and the Problem of the Package Deal*

### 4.1 *The Problem*

Consider a much discussed example (see, for example, Bennett 1980). Terror Bomber and Strategic Bomber both have as their goal promoting the war effort against Enemy. Both intend to pursue their goal by dropping bombs. Terror Bomber's plan is to bomb the school in Enemy's territory, thereby killing children of Enemy, terrorizing Enemy's population, and forcing Enemy to surrender. Strategic Bomber's plan is to bomb Enemy's munitions plant, thereby undermining Enemy's war effort. However, Strategic Bomber also knows that next to the munitions plant is a school and that when he bombs the plant, he will also kill the children inside th school. Strategic Bomber has worried a lot about this bad effect. But he ha concluded that this cost is outweighed by the contribution that would be made to the war effort by the destruction of the munitions plant.

Terror Bomber intends to drop the bombs, kill the children, terrorize the population, and thereby promote victory. In contrast, Strategic Bombe only intends to drop the bombs, destroy the munitions plant, and promot the war effort. While he knows that by bombing the plant he will be killing the children, he does not *intend* to kill them. Whereas killing the children is for Terror Bomber, and intended means to his end of victory, for Strategi Bomber it is only something he knows he will do by bombing the muni tions plant. Though Strategic Bomber has taken the children's deaths quit seriously in his deliberation, these deaths are for him only an expected side effect. Or so, anyway, it would seem.

But now consider a challenge to this distinction between intended and merely expected effects. We may address this challenge directly to Strategic Bomber: "In choosing to bomb the munitions plant, you knew yo would thereby kill the children. Indeed, you worried about this fact and took it seriously into account in your deliberation. In choosing to bom then, you have opted for a *package*, one that includes not only destr ing the plant and contributing to military victory but also killing the chi r The bombing is a *package deal*. So how can it be rational of you to intend only a proper part of this package?"

I call this *the problem of the package deal*.[1] This problem raises the following general question: Suppose an agent beleves her $A$-ing would result in bad effect $E$, seriously considers $E$ in her deliberation, and yet still goes o to make a choice in favor of her $A$-ing. How could it be rational of such a

agent not to intend to bring about *E*? We have supposed that a rational agent's intentions may fail to include expected effects she seriously considered in the deliberation that led to her choice. But how could this be, given that in deliberation and choice one is faced with a package deal?

To spell out the problem more explicitly, I will borrow from a plausible model of practical reasoning sketched by Wilfred Sellars in his essay "Thought and Action" (Sellars 1966). Sellars would see Strategic Bomber's reasoning as having three main stages. First, there is the stage in which he lays out the larger "scenarios" between which he must chose. These scenarios are to include those features of competing courses of action that are to be given serious consideration in his deliberation. In our example let us suppose that these scenarios will be

(S1) bomb munitions plant, destroy the plant, kill the children, and weaken Enemy

(S2) bomb the rural airport, destroy the airport, and weaken Enemy.

Second, Strategic Bomber will evaluate these scenarios "as wholes" and thereby arrive at a "complex intention." In our example this will be the complex intention to bomb and destroy the munitions plant, kill the children, and weaken Enemy. Finally, he will be led from this complex intention to the simpler intention to bomb the plant. And, though Sellars does not emphasize the point, it seems that Strategic Bomber will also be led to the simpler intention to kill the children.

This is a natural model. But so long as this is our model of Strategic Bomber's practical reasoning, we will have great difficulty explaining how he can rationally refrain from intending to kill the children. On the model, Strategic Bomber's initial conclusion of his practical reasoning will be a complex intention which includes his killing the children, and from which he seems to be as justified in reaching an intention to kill them as he is in reaching an intention to drop the bombs.

We can develop the point further by noting a quartet of principles to which this model of practical reasoning seems committed. These principles all concern the situation in which the practical reasoning of a rational agent has been successfully completed and has issued in an intention. There is, first, the idea that such practical reasoning should issue in a conclusion in favor of a scenario *taken as a whole*, where such scenarios include *all* factors given serious consideration in the reasoning. Strategic Bomber, for example, should draw a conclusion in favor of one of the total packages under consideration: (S1)-and-not-(S2), or (S2)-and-not-(S1). Second, there is the idea that this conclusion is a *practical* conclusion, one tied tightly to action. We may express this idea by saying that this practical conclusion is a *choice* of a total scenario. Strategic Bomber, then, should choose one of the total



packages under consideration; he cannot simply choose to drop the bombs. The third principle goes on to associate such a choice in favor of an overall scenario with the formation of a complex *intention* in favor of that overall scenario: Strategic Bomber must *intend* a total package. Finally, the fourth principle connects such an intention in favor of an overall scenario with intentions to perform each of the actions included within that scenario that the agent supposes to be within his control.

Here are slightly more precise statements of these principles:

> *Principle of the holistic conclusion of practical reasoning*: If I know that my *A*-ing will result in *E*, and I seriously consider this fact in my deliberation about whether to *A* and still go on to conclude in favor of *A*, then if I am rational, my reasoning will have issued in a conclusion in favor of an overall scenario that includes *both* my *A*-ing *and* my bringing about *E*. (For short, I will call this the *principle of holistic conclusion*.)
> *Principle of holistic choice*: The holistic conclusion (of practical reasoning) in favor of an overall senario is a *choice* of that scenario.
> *The choice-intention principle*: If on the basis of practical reasoning I choose to *A* and to *B* and to ..., then I *intend* to *A* and to *B* and to ....
> *Principle of intention division*: If I intend to *A* and to *B* and to ..., and I know that *A* and *B* are each within my control, then if I am rational, I will both intend to *A* and intend to *B*.[2]

Once we accept this quarter of principles, we are faced with the problem of the package deal. Strategic Bomber seriously considers the fact that by bombing, he will be killing the children. So the principles of holistic conclusion and holistic choice together require that (if rational) Strategic Bomber choose to bomb and to bring about the deaths of the children and to .... But then, by the choice-intention principle he must intend to bomb and to bring about the deaths of the children and to .... But Strategic Bomber knows that it is up to him whether to bomb and also that it is up to him whether to kill the children.[3] So by the principle of intention division Strategic Bomber, if rational, will intend to kill the children. So, contrary to our initial impression, there is no room here for a distinction between intended and merely expected upshots.

### 4.2 *Three Roles of Intention*

The problem of the package deal argues that, if rational, Strategic Bomber will intend to kill the children. The next step is to see why this conclusion should be rejected.

What is it to intend to kill the children? The problem of the package deal focuses on the prior reasoning on the basis of which intentions are *formed* But to understand what intentions are, we need also to look at the role

they play, once formed, in *further* reasoning and action. And here we can draw on our earlier discussion of these matters in sections 1–3.

Return to Terror Bomber. He really does intend to kill the children as a means to promoting military victory. As we have seen, this intention to kill the children will play two important roles in his further practical reasoning: it will pose problems for further means-end reasoning, and it will constrain his other intentions. To explore this second role, let us develop the example further. Suppose that after settling on his plan (but before his bombing run), Terror Bomber (who also commands a small battalion) considers ordering a certain troop movement. He sees that this troop movement would achieve certain military advantages. But then he notices that if the troops do move in this way, Enemy will become alarmed and evacuate the children, thereby undermining the terror-bombing mission. The option of moving his troops has an expected upshot (evacuation of the children) that is incompatible with an intended upshot of the bombing mission he already intends to engage in. But Terror Bomber's prior intention to terror-bomb, together with his beliefs, creates a screen of admissibility through which options must pass in later deliberation. And the option of moving the troops does not pass through this screen of admissibility, for it is incompatible with what Terror Bomber intends and believes. So Terror Bomber's prior intention to kill the children stands in the way of his forming a new intention to order the troop movement.

Now consider what happens when Terror Bomber executes his intention. Intentions are conduct *controllers*. This means that an intention to bring about some upshot will normally give rise to one's *endeavoring* to bring about that upshot. To endeavor to bring about some upshot is, in part, to guide one's conduct accordingly. Roughly, one is prepared to make adjustments in what one is doing in response to indications of one's success or failure in promoting that upshot. So Terror Bomber can be expected to guide his conduct in the direction of causing the deaths of the children. If in midair he learns they have moved to a different school, he will try to track them and take his bombs there. If he learns the building they are in is heavily reinforced, he may for that reason decide on a special kind of bomb. And so on.[4]

So, Terror Bomber's intention to kill the children will play a pair of characteristic roles as an input to his further practical reasoning and, when the time comes, it will lead to his endeavoring to kill them and so guiding his conduct in a way that aims at their death. What about Strategic Bomber's attitude toward his killing the children? It seems clear that his attitude toward killing them will *not* play such a trio of roles. Strategic Bomber will not see himself as being presented with a problem of how to kill the children. Nor will he be disposed to constrain his further intentions to fit with his killing them. If he were later to consider the troop movement just



described, and if he were to note the resulting likelihood of evacuation, this would *not* block for him the option of moving those troops. And finally, even once he is engaged in the bombing mission, he will not endeavor to kill the children. In the normal case this means he will not guide his activity by keeping track of the children and their deaths.[5] Rather, he will just keep track of the munitions plant and its destruction.

Further, there is good reason for Strategic Bomber to resist having an attitude toward killing the children that would play these three roles. Having such an attitude would not normally help him achieve those goals he wants to achieve, and it might well prevent him from achieving some goals he wants to achieve. For example, it might prevent him from considering the advantageous troop movement, given that option's expected incompatibility with his killing the children.

Since Strategic Bomber does not have an attitude toward killing the children that plays a trio of roles characteristic of intention, he does not intend to kill the children. And this may well be rational of him. So we should resist the pressure from the problem of the package deal to say that Strategic Bomber, if rational, will intend to kill the children.

So we are in the following situation. We have noted four principle that underlie the problem of the package deal. The inference from this quartet of principles to the conclusion that Strategic Bomber should intend to kill the children seems clearly valid. But our account of the functional roles characteristic of intention just as clearly indicates that this conclusion is to be rejected. This means we must reject at least one of these four principles. Our problem is, Which one?

### *4.3 Intention and Choice*

One might try rejecting intention division.[6] There is much to be said here, but for present purposes I just note that even without this principle a serious problem will remain. The remaining trio of principles still tells us that, if rational, Strategic Bomber will intend to kill the children *as a part of a larger intention* to bomb, and so on. But Strategic Bomber need not even have this larger, complex intention. Such a complex intention would play the trio of roles we have been focusing on. But it is clear, for example, that Strategic Bomber can be expected *not* to treat as inadmissible other options whose expected upshots are incompatible with his killing the children. Strategic Bomber may well later go ahead and seriously consider ordering the troop movement.

A second strategy would be to reject the principle of holistic conclusion. But, again, this does not seem promising to me. Granted, it would be too strong to require that a rational agent's practical conclusion include all expected effects of his action. Strategic Bomber expects that in going on his bombing run, he will be slightly heating up the wings of his aircraft. Yet

this might well not get into the conclusion of his practical reasoning. Since this effect does not matter to him in the least, Strategic Bomber may well not stop to notice it in his practical reasoning. But all that our principle of holistic conclusion requires is that the agent's conclusion include all expected upshots *that he has seriously considered in his deliberation.* What the principle requires is only a certain clearheadedness and intellectual honesty—an absence of "bad faith," if you will. Once I seriously consider *A*'s anticipated effect, *E*, in my deliberation about whether to *A*, I should see that the issue for my deliberation concerns a *complex* scenario, one that includes *A together with E.* If I am clearheaded and intellectually honest about this, my conclusion should concern this complex scenario, and not merely my *A*-ing simpliciter. My conclusion should concern the total package.

So we cannot solve our problem by rejecting either the principle of holistic conclusion or the principle of intention division. What then should we do?

Our problem arises from the apparent conflict between backward-looking and forward-looking pressures on what we intend. On the one hand, intentions are typically grounded in prior deliberation. Once deliberation enters the picture, however, plausible standards of clearheadedness and intellectual honesty—standards expressed in the principle of holistic conclusion—are engaged. This leads to pressure *for* practical conclusions to be holistic and so, it seems, for intentions to be holistic. This pressure is backward-looking, for it is grounded in the connection between intention and prior deliberation. On the other hand, intentions, once formed, play a trio of characteristic roles in further reasoning and action. This is true not only of a relatively simple intention to drop bombs but also of a more complex intention to drop bombs and (thereby) kill the children. Since Strategic Bomber, for example, can rationally refrain from having an attitude toward killing the children that plays such roles, there is pressure *against* forcing intentions to be holistic. This pressure is forward-looking, for it depends on the roles of intentions in guiding further reasoning and action. The problem of the package deal is the problem of how to reconcile these conflicting pressures. The solution is to deny that these pressures apply to the same thing. I proceed to explain.

The problem of the package deal depends on identifying (or, anyway, linking very tightly) (1) the conclusion of practical reasoning that is subject to pressures for holism (pressures based on standards of clearheadedness and intellectual honesty in reasoning) and (2) the intention in which practical reasoning issues. This identification derives from the combination of the principle of holistic choice and the choice-intention principle. The former identifies (1) with a holistic choice; the latter connects such a choice with (2). The rejection of either of these principles would undermine the identification of (1) with (2). This would allow us to say that the pressure



for holism, though it applies to certain conclusions of practical reasoning (that is, (1)), does not apply to the intentions in which the practical reasoning issues (that is, (2)). And in this way we could block the problem of the package deal.

I propose, then, that we challenge either the principle of holistic choice or the choice-intention principle. But which one? Though I cannot argue the point here, it seems clear to me that we should reject the choice-intention principle. We should distinguish what is chosen on the basis of practical reasoning from what is intended. Choice and intention are differently affected by standards of good reasoning, on the one hand, and concerns with further reasoning and action, on the other.

Return to Strategic Bomber. He is obliged by a plausible principle of good reasoning to include the children's deaths in the total package that he chooses on the basis of his *prior* reasoning. But it does not follow that his attitude toward these deaths must, if he is rational, play the roles in *further* reasoning and action that are characteristic of intention. The demand to include the children's deaths in what is chosen comes from a demand for clearheadedness in one's reasoning and choice—a demand to confront clearly and honestly the important features of what one will be doing in plumping as one does. It is this demand that leads to pressure for choice to be holistic. But this demand does not force one's *intentions* to be holistic; for one's intentions are tied not only to prior deliberation but also to a trio of roles concerning further reasoning and action. Nothing in the ideal of clearheaded reasoning forces one to have the dispositions concerning *further* reasoning and action that would be characteristic of holistic intention. Though clearheadedness obliges Strategic Bomber to choose (among other things) to kill the children, it does not oblige him later to screen out options incompatible with his killing them, or to endeavor to kill them. So choice and intention can diverge.

It is natural to assume that whatever one choose one thereby intends (see, for example, Aune 1977, 115). But reflection on the problem of the package deal argues otherwise. What one chooses is constrained by holistic pressures—pressures grounded in standards of clearheadedness in reasoning—in a way in which what one intends is not. And what one intends is tied to further reasoning and action in a way in which what one chooses need not be.

Our model of the roles of intention in practical reasoning sees intentions as both characteristic inputs and outputs of such reasoning. What we have seen here is that what intentions are an output of present practical reasoning is constrained in part by the roles of such intentions as inputs to further practical reasoning.

Of course, having rejected the choice-intention principle, we will need to put something in its place. One's choice in favor of an overall scenario will

involve one's coming to have some intentions or others—intentions that will guide further reasoning and action. This is what distinguishes a choice of a scenario from a mere preference or positive evaluation in its favor. A full theory will need to say more about *which* intentions a rational agent will thereby come to have; but I cannot pursue this matter further here. Suffice it to say that an acceptable treatment of the relation between intention, choice, and practical reasoning cannot just assume that one intends all of what one chooses on the basis of practical reasoning.

## 5 *Concluding Remarks*

This concludes my brief sketch of aspects of my approach to saying what intention is and to identifying some of the major roles of intention in the intelligent activity of limited agents like us. Intentions are relatively stable attitudes that function as inputs to further practical reasoning in accordance with the two-level model of practical reasoning I have outlined. Intentions will be outputs of practical reasoning, but in a way that comports with the complexity of the relation between choice and intention. As a conduct-controlling pro-attitude, intention is intimately related to endeavoring and action.[7]

What about our original trilemma? Future-directed intentions influence later conduct by way of their influence on intervening practical reasoning and the formation of derivative intentions, by way of their stability, and by way of their tendency to control conduct when the time come. This is not action at a distance; but it is also not irrevocability. These systematic ways in which future-directed intentions shape later deliberation and action are central to the functioning of limited but intelligent agents like us, especially in the pursuit of coordination. So there is little reason to worry that the formation of such intentions is a waste of time. Our model of the role of intention in limited rational agency seems clearly to avoid the horns of our trilemma and so to pass at least this test of adequacy for an account of the nature of intention.

## *Notes*

1. In the discussion to follow I benefited from Gilbert Harman's examination of a related but different problem in Harman 1983. I explain how my problem differs from Harman's, and why I reject important features of his solution, in Bratman 1987, chapter 10, from which the present discussion is taken.
2. Note that this last principle is rather limited. For example, it does *not* say that if I intend to X and I know that it is necessary that if I do X I do Y, then I intend to Y. For the principle of intention division to apply, I must know that X and Y are, each of them, within my control. So, on the present principle, I might intend to belch softly and yet (rationally) not intend to belch, even though I know that it is necessary for me to belch if I am to belch softly. This might happen if I thought that it was not up to me whether I belched, but only up to me whether, given that I was going to belch, I belched softly c loudly. (The example is due to Gilbert Harman, in correspondence.)
3. Granted, he also knows that it is *not* in his power to bomb *without* killing the childre But that does not affect the present point.
4. Of course, this all depends on his retaining his beliefs about the connection between h killing the children and his promoting military victory. If he finds out, in midair, that war is over, or that the children in the school are prisoners Enemy would like to killed, he will not continue to guide his conduct by the children's death.
5. I say "in the normal case," for there are cases in which Strategic Bomber might guide h activity by keeping track of the children and their deaths and yet still not endeavor to k them. Suppose he wasn't dropping a single bunch of bombs but was launching missil one at a time, aimed at the munitions plant. However, the only way he has of knowir whether he has hit the plant is to listen to Enemy radio for an announcement of t deaths of the children. After each missile is launched, he waits for a radio announceme In its absence he launches yet another missile, and so on until he hears of the child deaths or somehow comes to change his opinion of the link between the destructior the munitions plant and the deaths of the children. Still, he does not launch his missile *order to* kill the children. Nor does he engage in means-end reasoning concerned wi how to kill the children, or screen his other options for their compatibility with his killi them. So he still does not intend to kill them.
6. This is, in effect, what Chisholm (1976) suggests; see especially chapter 2, pages 74–
7. Once we see intention in this way, how should we understand the relation betw intending to act and our classification of actions as done intentionally or with a cer intention? I discuss these matters in Bratman 1987, chapters 8 and 9.



## *References*


Anscombe, G. E. M. (1963). *Intention.* 2nd ed. Ithaca, NY: Cornell University Press.

Aune, Bruce (1977). *Reason and action.* Dordrecht, Holland: D. Reidel.

Bennett, Jonathan (1980). Morality and consequences. In Sterling M. McMurrin, ed., *Tanner lectures on human values.* Cambridge: Cambridge University Press.

Bratman, Michael (1983). Castañeda's theory of thought and action. In James E. Tomber ed., *Agent, language, and the structure of the world: Essays presented to Hector-Neri Castañ with his replies.* Indianapolis, IN: Hackett.

Bratman, Michael (1985). Davidson's theory of intention. In Bruce Vermazen and Merri Hintikka, eds., *Essays on Davidson: Actions and events.* New York: Oxford Unive Press. Reprinted with an added appendix in Ernest LePore and Brian McLaughlin, (1985). *Actions and events: Perspectives on the philosophy of Donald Davidson* xfo Basil Blackwell.

Bratman, Michael E. (1987). *Intention, plans, and practical reason.* Cambridge, MA: Harv University Press.

Castañeda, Hector-Neri (1975). *Thinking and doing.* Dordrecht, Holland: D. Riedel.

Chisholm, Roderick M. (1976). *Person and object.* La Salle, IL: Open Court.

Davidson, Donald (1980). *Essays on actions and events.* New York: Oxford University Pr

Grice, H. P. (1974–75). Method in philosophical psychology: From the banal to the biz *Proceedings and Addresses of the American Philosophical Association* 48, 23–53.

Harman, Gilbert (1983). Rational action and the extent of intentions. *Social Theory Practice* 9, 123–41. Revised version published as chapter 9 in Gilbert Harman (19 *Change in view.* Cambridge, MA: MIT Press.

Jeffrey, Richard (1983). *The logic of decision.* Chicago: University of Chicago Press.

Sellars, Wilfred (1966). Thought and action. In Keith Lehrer, ed., *Freedom and determini* New York: Random House.